

## **Supporting Information**

### **Metatranscriptomic analysis of microbial community in an ocean-front deep subsurface hot spring reveals novel small RNAs and type-specific tRNA degradation**

Shinnosuke Murakami<sup>1,2</sup>, Kosuke Fujishima<sup>1</sup>, Masaru Tomita<sup>1,2,3</sup>, Akio Kanai<sup>1,2,3\*</sup>

<sup>1</sup>Institute for Advanced Biosciences, Keio University, Tsuruoka, 997-0017, Japan

<sup>2</sup>Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, 252-8520, Japan

<sup>3</sup>Faculty of Environment and Information Studies, Keio University, Fujisawa, 252-8520, Japan

\*Corresponding author:

Akio Kanai, PhD

Institute for Advanced Biosciences, Keio University,

Tsuruoka, Yamagata 997-0017, Japan

Tel: +81-235-29-0524; Fax: +81-235-29-0525; E-mail: akio@sfc.keio.ac.jp

**Supplementary Table S1. Information and accession numbers for the 16S rRNA sequences determined in this study.**

**Supplementary Table S2. Information and BLAST results for SURFYs.** “–” indicates that no similar sequence was detected with a BLAST search of the whole NCBI nucleotide database.

**Figure S1. PCR amplification of microbial 16S rRNA genes.**

16S rRNA genes were amplified using either (A) a recently reported archaeal primer set (340F and 1000R), (B) the originally designed archaeal primer set (340F2 and 932R), or (C) the universal bacterial primer set (341F and 907R). Genomic DNA isolated from the Yunohama hot spring water was used as the template for all PCRs. The amplified PCR products (P) were visualized by ethidium bromide staining after 1.5% agarose gel electrophoresis with the molecular marker φX174/*Hae*III (M). Arrowheads indicated the positions of specific PCR products.

**Figure S2. Gel electrophoresis of the cloned small RNA cDNA library used for deep sequencing.**

A cDNA library of Yunohama-derived small RNAs (<200 nt) was constructed using the Small RNA Cloning Kit (Takara, Japan) and was visualized by 10%–20% polyacrylamide gel electrophoresis (S) together with a 20-bp marker ladder (M).

**Figure S3. Phylogenetic positions of the archaeal 16S rRNA gene sequences isolated from the Yunohama hot spring water.**

An archaeal phylogenetic tree was constructed for 318 archaeal 16S rRNA gene sequences, including 13 nonredundant Yunohama-derived sequences, using a bacterial 16S rRNA as the outgroup (gray). The Yunohama-derived sequences were obtained by PCR cloning using either the recently reported archaeal primer set (orange) or the originally designed primer set (magenta). The phylogenetic tree was constructed based on the neighbor-joining method, with 100 bootstrap iterations (bootstrap values > 70 are shown on each branch). Branch colors represent the different phyla or groups containing sequences derived from the Yunohama hot spring. The numbers of sequenced reads are shown in brackets.

**Figure S4. Comparison of the intervening repeat sequences found in the 16S rRNA genes of the Yunohama ARMANs.**

Similar intervening repeat sequences (#1, #2, #3, and #4) were found in two distinct regions (A or B) of the 16S rRNA gene sequences from the Yunohama ARMAN. The arrows indicate the nucleotide positions in 16S rRNA gene sequence of ARMAN-2. Comparisons of each intervening repeat sequence (C). The sequences were aligned using ClustalW 2.0. Gaps (–) were inserted to maximize the numbers of nucleotide matches. The aligned sequences in C were also manually improved to maximize the number of nucleotide matches.

**Figure S5. Phylogenetic positions of bacterial 16S rRNA sequences isolated from the Yunohama hot spring water.**

A bacterial phylogenetic tree was constructed for 331 bacterial 16S rRNA sequences, including 15 nonredundant Yunohama-derived sequences (magenta), using an archaeal 16S rRNA as the outgroup (gray). The method of phylogenetic tree construction and other information are the same as those described in Figure S3.

**Figure S6. Correlation analysis of read numbers for the mature tRNAs and their fragments.**

A scatter plot of mature tRNAs (numbers of reads of mature tRNA sequences) and tRNA fragments (numbers of reads of partial tRNA sequences) is shown ( $N = 5,448$ ). The correlation coefficient is  $R = 0.15$ .

**Figure S7. Most frequent cleavage sites for each tRNA in the anticodon table.**

The sites most frequently cleaved in the anticodons of the tRNAs are shown. “V” indicates variable stem/loop regions. Two types of cleavage sites are colored: 34/35 (blue) and outside the anticodon region (magenta). No fragmented tRNAs with anticodons corresponding to blank cells (white) were isolated. Because the number of fragmented tRNAs with anticodons corresponding to “–” was low, the rankings of these cleavage sites were not calculated.

**Supplementary Table S1: Information and accession numbers of 16S rDNAs determined in this study**

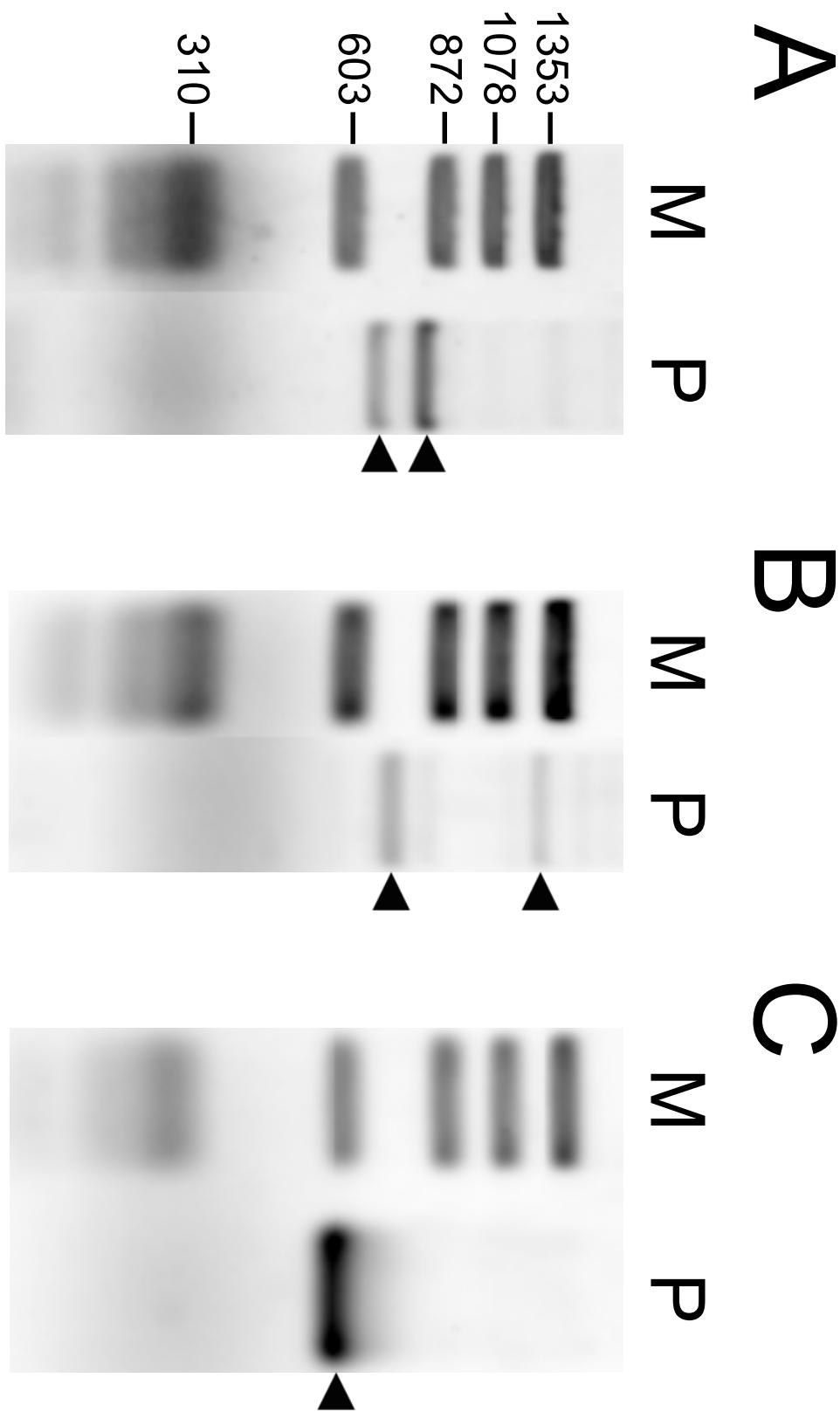
	Clone #	Length (nt)	Number of related clones	Putative taxonomy	DDBJ/EMBL/GenBank accessions
Archaea (340F-1000R)	A-110	784	53	ARMAN	AB679512
	A-115	646	14	Methanococcales	AB679514
	A-127	738	5	ARMAN	AB665425
	A-121	641	5	SAGMCG	AB679517
	A-120	645	3	Methanococcales	AB679516
	A-112	640	3	MCG	AB679513
	A-116	640	2	MCG	AB679515
	A-147	640	2	MCG	AB679518
Archaea (340F2-932R)	A-14	657	52	ARMAN	AB679519
	A-9	1187	6	HWCG I	AB665420
	A-52	611	3	ARMAN	AB683043
	A-32	507	1	AAG	AB665423
	A-54	513	1	FSCG	AB665424
	A-12	676	1	Marine Group	AB665421
	A-17	515	1	Thermoplasmatales	AB665422
Bacteria	B-73	548	48	β-Proteobacteria	AB665436
	B-65	548	20	β-Proteobacteria	AB665434
	B-35	549	7	β-Proteobacteria	AB665431
	B-10	542	5	Bacteroidetes	AB665427
	B-15	523	3	Aquificae	AB665428
	B-32	541	1	Bacteroidetes	AB665430
	B-38	542	1	Bacteroidetes	AB665432
	B-26	543	1	Bacteroidetes	AB665429
	B-71	541	1	<i>Deinococcus/Thermus</i>	AB665435
	B-3	551	1	δ-Proteobacteria	AB665426
	B-47	525	1	δ-Proteobacteria	AB665433
	B-75	550	1	δ-Proteobacteria	AB665437
	B-83	552	1	δ-Proteobacteria	AB665438
	B-96	550	1	γ-Proteobacteria	AB665440
	B-91	540	1	Unknown	AB665439

**Supplementary Table S2: Information and BLAST results for SURFYs**

DDBJ/EMBL/ GenBank accessions	Small RNA Information					BLAST Result		
	Name	Reads	Length (nt)	GC (%)	$\Delta G$ (kcal/mol)	Gene	Identity (%)	Coverage (%)
AB665588	SURFY 01	19	38	60	-6.7	-	-	-
AB665589	02	17	101	57	-50.2	-	-	-
AB665590	03	14	68	58	-21.6	EF-Tu	91	83 <i>Nitrosococcus halophilus</i>
AB665591	04	14	27	51	-0.8	-	-	-
AB665592	05	13	62	59	-34.8	-	-	-
AB665593	06	12	43	48	-10.6	-	-	-
AB665594	07	12	46	58	-14.4	-	-	-
AB665595	08	12	71	67	-32.7	-	-	-
AB665596	09	11	75	56	-30.1	-	-	-
AB665597	10	11	38	60	-6.3	-	-	-
AB665598	11	11	31	54	-2.9	-	-	-
AB665599	12	10	98	61	-33.5	-	-	-
AB665600	13	10	49	42	-6.4	-	-	-
AB665601	14	10	62	61	-13.6	-	-	-
AB665602	15	10	85	76	-35.9	-	-	-

“-” indicates that no similar sequence was detected with a BLAST search against the whole NCBI nucleotide database.

**Fig.S1**



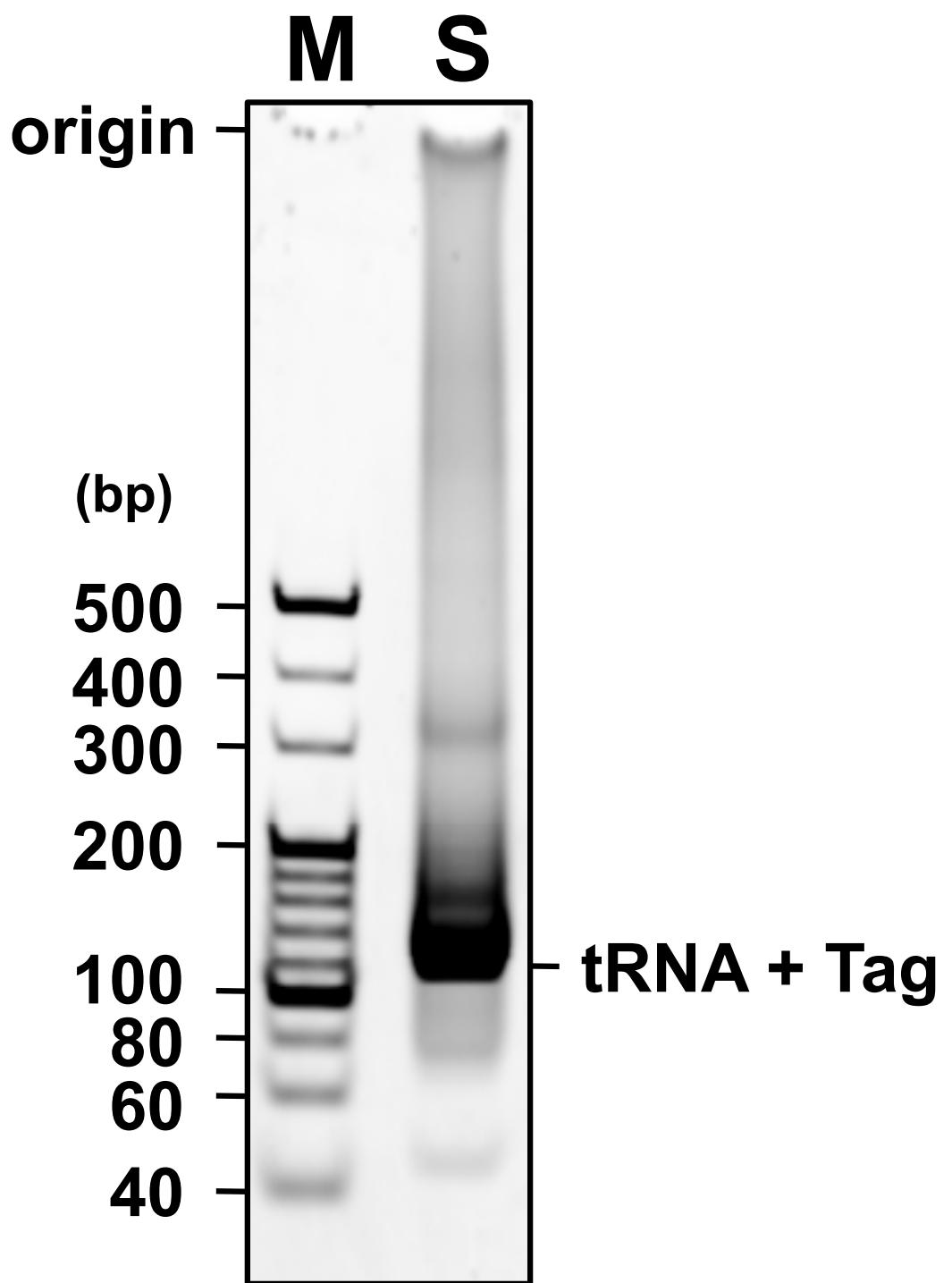


Fig.S2

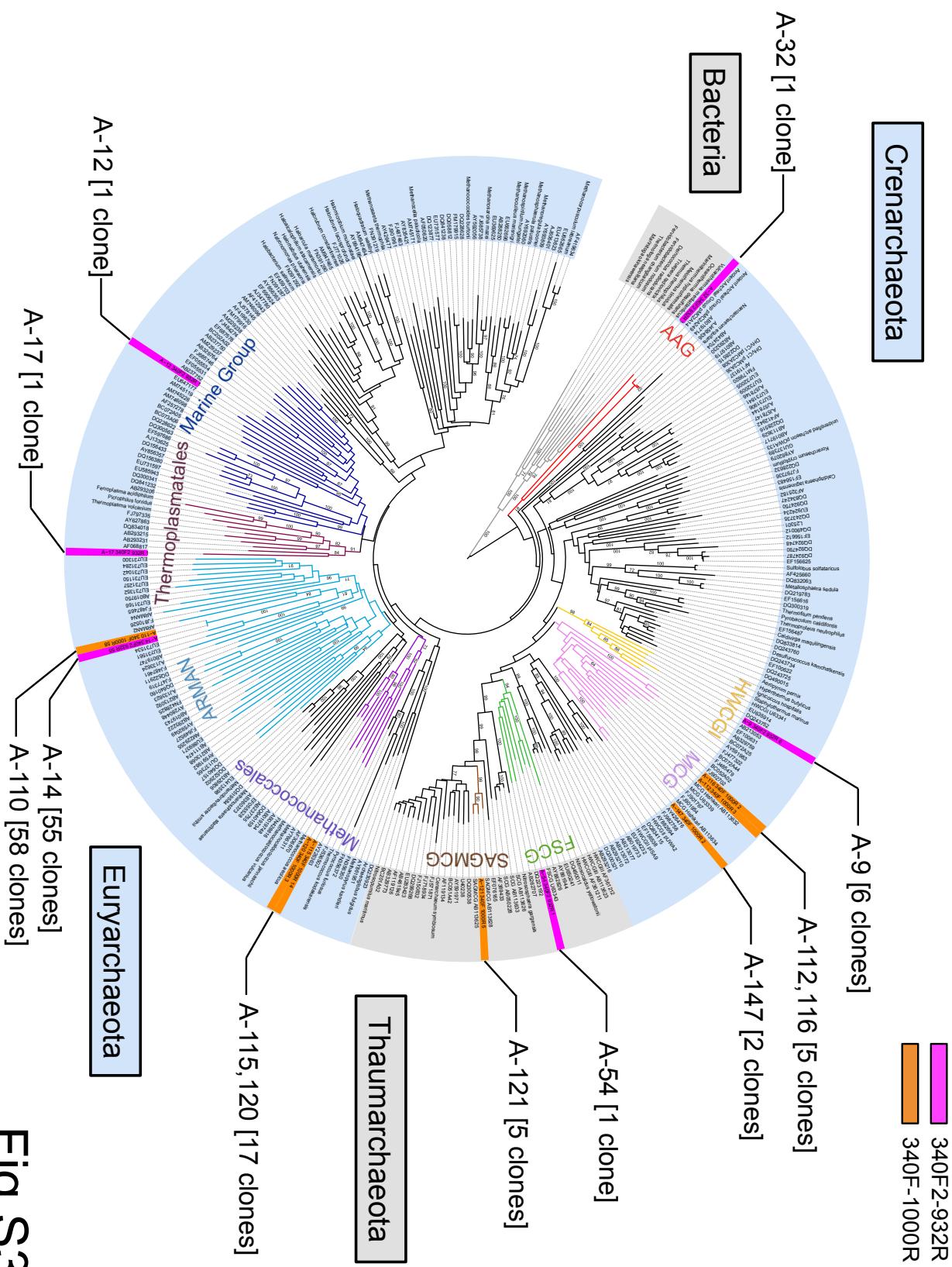
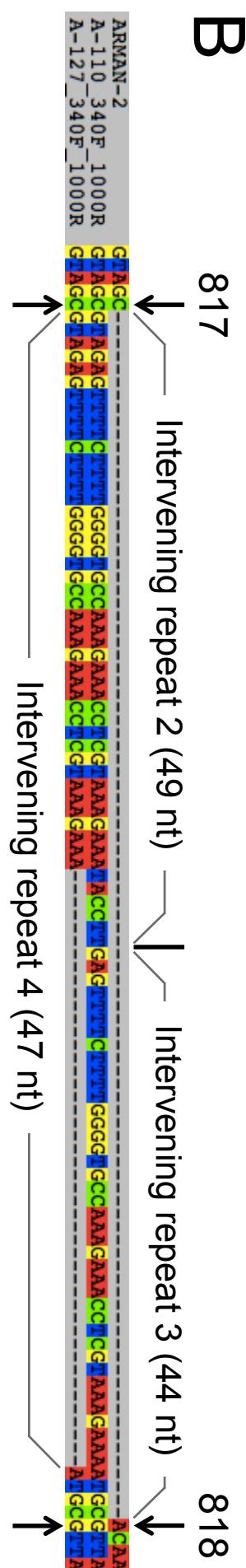
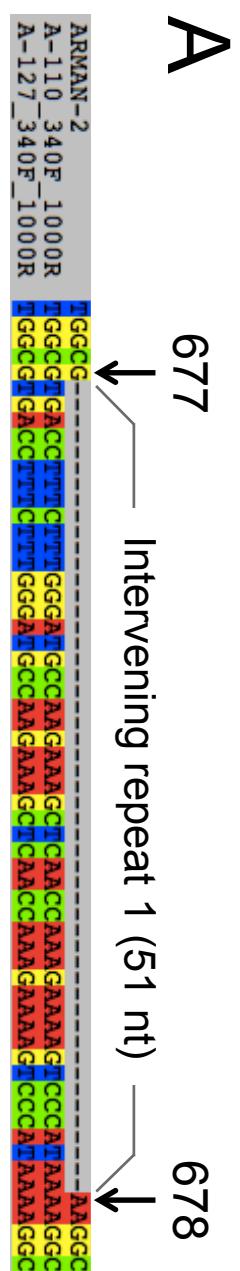


Fig. S3



Intervening repeat 1	-T-GACCTTCCTT-GGGATGCC-AAGANAGCTAACCAAGAAGTCCTATAA
Intervening repeat 2	G-TAGAGTTCTTCTTGGGTGCCAAGAACCTCGT--AAGAA-TACCTT
Intervening repeat 3	-GAGTTTCTTGGGTGCCAAGAACCTCGT--AAGAA-ATGCC
Intervening repeat 4	G-TAGAGTTCTTGGGTGCCAAGAACCTCGT--AAGAA-AIGC

**Fig.S4**

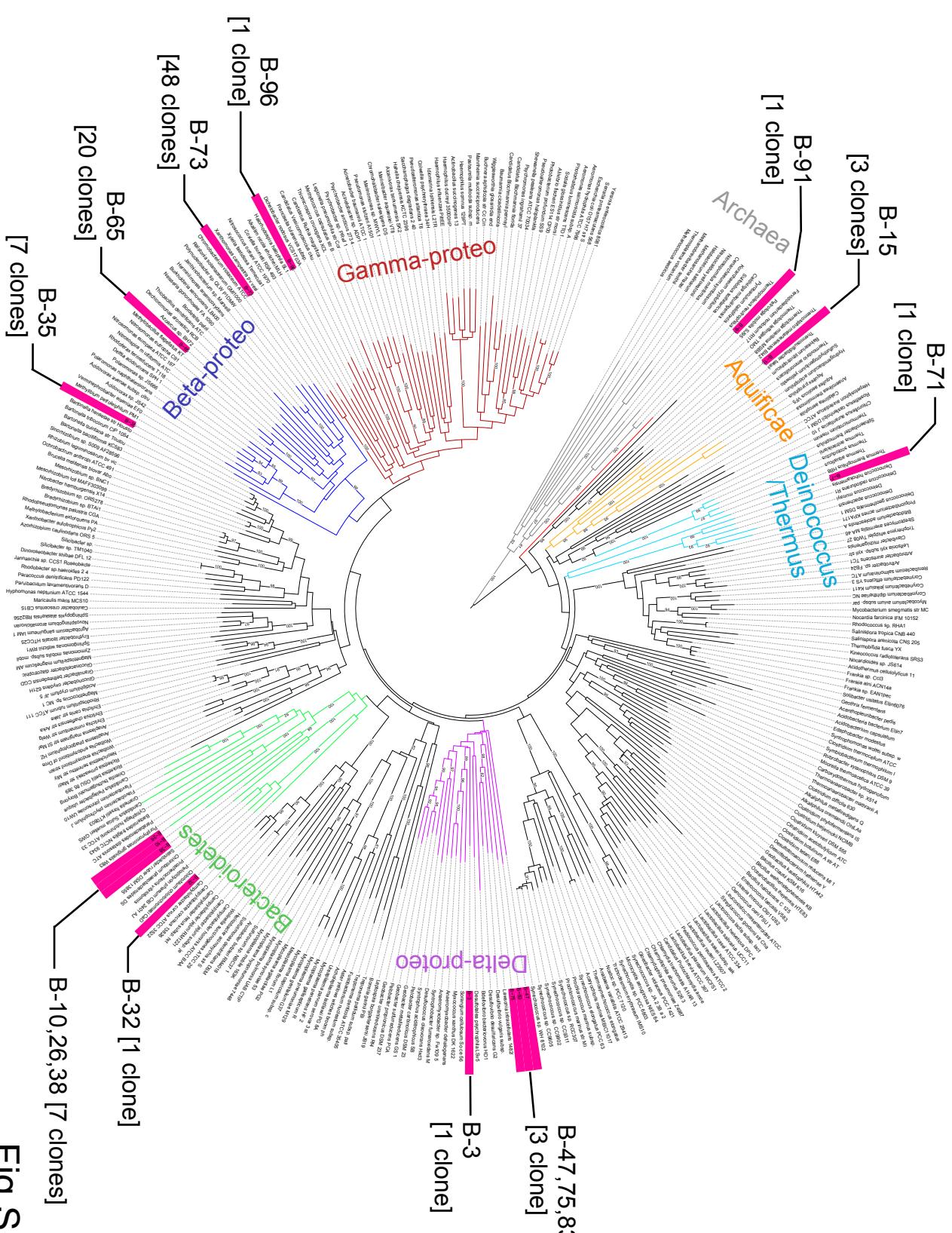
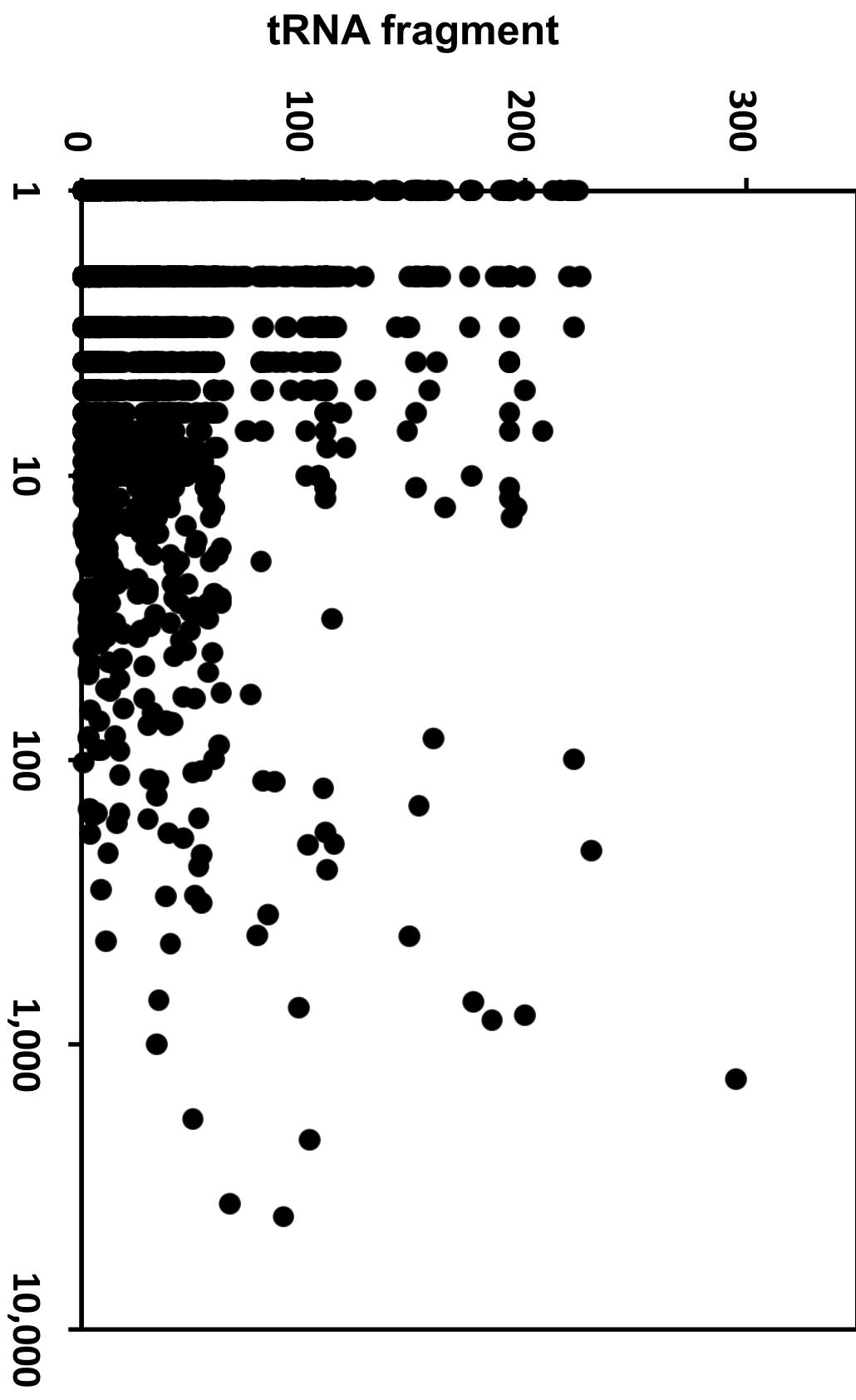


Fig.S5

**Fig.S6**

Mature tRNA



		2nd base of anticodon					
		A	G	T	C		
1st base of anticodon	A						A
		AAG	-		ATG 37/38	ACG 34/35	G
							T
			AGC 35/36				C
	G	GAA 34/35	GGA 35/36	GTA -	GCA 36/37	A	
		GAG 34/35	GGG 36/37	GTG 36/37	GCG 34/35	G	
		GAT 15/16	GGT 35/36	GTT 38/39	GCT 34/35	T	
		GAC 34/35	GGC 34/35	GTC 35/36	GCC 34/35	C	
	T		TGA 35/36	TTA 35/36			A
		TAG V	TGG 36/37	TTG 36/37	TCG 37/38	G	
			TGT 15/16	TTT 15/16	TCT 57/58	T	
		TAC 15/16	TGC 35/36	TTC 35/36	TCC 35/36	C	
	C	CAA -	CGA V	CTA V	CCA 25/26	A	
		CAG 16/17	CGG 35/36	CTG 37/38	CCG 15/16	G	
		CAT 36/37	CGT 35/36	CTT 19/20,57/58	CCT 15/16	T	
		CAC 15/16	CGC 35/36	CTC 45/46	CCC 55/56	C	

Fig.S7